

Special Issues Analysis Center, Task Order Report, March 28, 1994

AN EXAMINATION OF ASSESSMENT OF LIMITED ENGLISH PROFICIENT STUDENTS

Task Order D070 Report

Prepared by:
Annette M. Zehler
Paul J. Hopstock
Howard L. Fleischman
Cheryl Greniuk

Submitted by:
Special Issues Analysis Center
Development Associates, Inc.
1730 North Lynn Street
Arlington, VA 22209

March 28, 1994

Note: Every attempt has been made to maintain the integrity of the printed text. In some cases, figures and tables have been reconstructed within the constraints of the electronic environment.

TABLE OF CONTENTS

[I. Introduction](#)

- A. Background
- B. Purpose of the Report

[II. Issues Related to the Use of Assessment](#)

- A. Inclusion of LEP Students in Effective Instruction: Is Assessment Reform Also Instructional Reform?
- B. Assessment in Support of Educational Decision-Making for LEP Students: How Should Assessment be Used to Determine Eligibility and Placement?
- C. Assessment for Accountability: To What Extent Should LEP Students Be Included

[III. The Purposes of Assessment](#)

[IV. Assessment Practices Associated with Successful Programs](#)

[V. A Review of English Language Proficiency Tests](#)

- A. Examination and Comparison of the Language Proficiency Tests
- B. Theoretical Bases of the Tests
- C. Issues Related to Reliability and Validity

[VI. A Review of Most Frequently Used Standardized Achievement Tests](#)

[VII. Issues in the Use of Standardized Tests](#)

- A. Criticisms of Standardized Achievement Tests
- B. Concerns Regarding the Language of Administration in Standardized Testing

VIII. Alternative Assessment

- A. Definition of Alternative Assessment
- B. Motivations for a Shift Toward Alternative Forms of Assessment
- C. Issues in the Development of Alternative Assessment Approaches
- D. Issues in the Implementation of Alternative Assessment
- E. Future Directions
- F. Summary

IX. Conclusions and Recommendations

- A. Assessment in General
- B. Assessment of LEP Students for Instructional Purposes
- C. Assessment of LEP Students for Identification and Placement Purposes
- D. Assessment of LEP Students for Evaluation and Accountability Purposes

Appendix A: References

LIST OF TABLES

1. [The Importance of Various Assessment Characteristics for Different Assessment Purposes: Programs with Primarily Native Language Instruction](#)
2. [Comparisons of Six Most Frequently Used Language Proficiency Tests](#)
3. [Six Most Frequently Used Language Proficiency Tests: Comparison of Administrative Procedures](#)
4. [Comparisons of Five Most Frequently Used Achievement Tests](#)

I. INTRODUCTION

A. Background

The publication of *A Nation at Risk* (1983) focused attention on the need to restructure the practices of our educational system. Findings that U.S. students lag far behind students in other countries prompted a broadly shared concern regarding the level of skills of students exiting our educational institutions and entering the workforce. In response to this concern, the decade of the 80's produced substantial efforts toward educational reform. The definition of the six National Education Goals in 1989 focused these efforts by identifying specific objectives for the nation to achieve by the year 2000. Now, as we head well into the 90's, more recent legislative initiatives, such as reflected in the proposed reauthorization of the Elementary and Secondary Education Act (ESEA), emphasize the need for full opportunity for every child--including limited English proficient (LEP) children and others with special needs--to achieve at high levels (Garcia, 1994). These expectations for achievement are being made more specific as a variety of groups and organizations carry out efforts to define national standards in a range of content areas.

The principle of inclusion of all children in effective instruction is an important element in the recently proposed ESEA legislation. In the case of LEP students, the need to ensure their full participation in

effective instruction is seen in their growing numbers. Data from the 1990 Census show that about 16 percent of the 91.7 million households in the U.S. are households in which a language other than English is used; about half of the language minority households include children under 18 years of age (Waggoner, 1994). Recent data show that LEP students in grades K-12 now number approximately 2.3 million (Fleischman and Hopstock, 1993)--an increase of one million since the last such estimate made in fall of 1983--and represent 5.5 percent of all students in grades K-12 in public schools across the country. Clearly, as we move closer to the year 2000, the debate on how to best promote high achievement for all students must address questions specific to language minority and LEP students (LaCelle-Peterson and Rivera, 1994).

There has been considerable recent attention by educational policy-makers and practitioners to the issues of why, when, and how limited English proficient (LEP) students should be assessed. Two particular trends in the educational reform movement have spurred this attention. One trend has involved the reexamination of existing educational assessment measures and practices in terms of their relationship to effective instruction, their fairness to minority populations, and their use in making decisions about specific students. The second trend has been related to the development of educational standards for students, leading to questions of when and how standards and related assessments should be applied to LEP students.

B. Purpose of the Report

The purpose of this report is to examine assessment instruments and practices and to review issues related to assessment of LEP students. It is critical to examine these issues now, as major steps are being taken toward: (1) the redesign of assessment approaches; and (2) definition of standards and new national assessment systems. Given their increasingly large representation in schools today, issues related to language minority and LEP students should be included early on in the efforts to define new assessments, especially those related to national standards (DeAvila, 1994).

The report begins by considering three key issues related to the assessment of LEP students, and then discusses the various purposes of assessment and the important concerns related to those purposes. The next chapters present examinations and comparisons of the most frequently used language proficiency tests and standardized academic achievement tests. These are followed by a discussion of more recent directions in assessment and the implications of these newer efforts for LEP students. The final chapter presents conclusions and recommendations for the assessment of LEP students.

II. ISSUES RELATED TO THE USE OF ASSESSMENT

Inclusion of LEP students in assessment is currently being discussed within three key sets of issues: (1) inclusion of LEP students in assessment as part of effective and challenging instruction that promotes the development of high levels of achievement; (2) appropriate inclusion of LEP students within assessments that determine student placement or eligibility for services; and (3) inclusion of LEP students within assessments that examine accountability (of programs, schools, districts, etc.). Overall, the issue is equity for LEP students.

A. Inclusion of LEP Students in Effective Instruction: Is Assessment Reform Also Instructional Reform?

A first question regarding assessment concerns the relationship between assessment and instruction. One of the criticisms of current standardized tests (discussed further in Section VII below) is that tests begin to

drive instruction and change the nature of the curriculum as teachers "teach to the test". This has been an unwanted outcome of the standardized tests, resulting in lower levels of instruction for students.

For this reason, assessment reform is seen by many as critical to instruction, due to the very powerful effect that assessment has on the nature of instruction (Darling-Hammond, 1994). According to this view, if assessment methods themselves are directed toward the same skills that should be developed, then quality instruction is more likely to occur, even if teachers "teach the test." Darling-Hammond (1994) defines this as "consequential validity." This refers to the positive impact of an assessment tool on the teaching and learning process. In this way, "teaching to the test" is turned toward more positive outcomes.

The use of newer forms of assessment (e.g., performance assessment, alternative assessment) is expected to promote more effective classroom instruction. This is instruction that is focused on more authentic student-centered learning, development of higher order thinking skills, and constructivist approaches to learning. These types of instructional approaches, which are consistent with use of alternative assessments, are expected to be effective for all students. However, they should be especially beneficial to LEP students. For example, instructional tasks and contexts described as "authentic" are also the types of settings in which LEP students will experience more opportunities for meaningful language use, for contextual support for learning, and for participation in instruction and assessment through a variety of formats. The use of alternative assessment is viewed as a potentially strong and effective agent of change toward these effective instructional approaches.

The use of alternative assessments will not resolve all problems, however. If alternative assessments are implemented, then there are additional issues to consider. For example, it is very likely that the use of alternative assessments will require even more language use by students in responding to the assessment tasks (Secada, 1994). Will LEP students therefore be placed at a greater disadvantage in alternative assessment tasks? An additional question concerns the linkage between the assessments used and the curriculum received by LEP students. For example, if LEP students do not receive instruction that provides for student-directed instructional activities, a focus on higher order cognitive tasks, or opportunities for hands-on problem-solving activities, then use of performance assessment tasks for LEP students will not be as appropriate and outcomes obtained will be less valid. That is, performance assessment tasks will not be fair for LEP students who have not been exposed to these types of instructional activities. The relationship between the contexts and activities presented in the assessment tasks and those the student typically encounters in the classroom can be a critical one.

B. Assessment in Support of Educational Decision-Making for LEP Students: How Should Assessment Be Used to Determine Eligibility and Placement?

Assessment is used to support decisions regarding the education of students. For LEP students, assessment first involves identification as a LEP student who is eligible for special language-related services. Assessment also is used to determine student placement, to determine eligibility for other special programs, to reclassify students from LEP status, and to make decisions regarding promotion or graduation. In these uses, the important decisions that are made affect subsequent opportunities for students.

Initial assessment that identifies a student as LEP provides important access to special services. However, at the same time, assessment and classification as LEP may also limit the student in terms of future opportunities and unfairly exclude LEP students from other opportunities, such as for advanced or other special instruction. Research on use of assessment has shown that LEP students are often assessed and inappropriately placed within special education, where they are more likely to receive low level instruction and less challenging content. Similarly, the use of single assessment measures that are inappropriate to LEP

students may be the only means of qualifying for gifted and talented instruction or other instructional opportunities. Decisions regarding promotion or graduation are similarly critical to future opportunities for LEP students. Lack of a high school diploma will make finding employment more difficult, yet promotion or graduation decisions may be made based on assessment measures that are not appropriate to the LEP student.

It is preferable to base decisions regarding a student's education on multiple measures. Ortiz and Wilkinson (1990) have demonstrated the effect of using multiple approaches and more appropriate assessment methods. In their research, they show that with better "prereferral intervention," inappropriate assignment of LEP students to special education can be decreased. Even with multiple measures, however, there are a number of questions which need to be addressed: What specific tests are most appropriate for use? To what extent should native language testing be used? Who makes decisions using the multiple measures, and what criteria should be used? How should the various measures be weighted in making placement decisions? Without appropriate responses to these issues, multiple measures may not lead to more valid decision-making.

C. Assessment for Accountability: To What Extent Should LEP Students Be Included?

The large-scale assessment programs that have increased over recent years are designed to serve purposes of program evaluation or school/district accountability. In these, there have been serious questions raised regarding the appropriateness of testing for LEP students. There is first of all a concern for the LEP students, who are being required to take a test for which their level of English is not sufficient. Yet, it may not be any more beneficial to the students to be tested in their native language if that has not been the language of instruction.

A second issue concerns the LEP students' access to the same curriculum received by non-LEP students. For example, instruction received by LEP students may be slower paced or changed in other ways and LEP students may not have access to the same content courses (Minicucci and Olsen, 1992). If the curriculum which is assumed for the test has not been received by the LEP students, then the outcomes obtained for LEP students will not provide for valid comparisons.

As a third issue, there are criticisms that the tests are not fair for LEP students in that they are culturally biased in favor of mainstream, middle class values and experiences. This criticism is made both in reference to test item content and to the structure of the assessments themselves (e.g., the use of timed tests in which students must understand that their best performance within a limited period of time is required).

A fourth issue concerns the legacy of misuse of testing with language minority populations (e.g., Hakuta, 1986). Past experiences in which test results from language minority students have been inappropriately interpreted has generated a mistrust of large-scale testing. Thus, the issue of inclusion of LEP students within large-scale testing has been controversial. At the program and classroom level, in particular, educators who work with LEP students are concerned that inclusion of their LEP students in tests will not benefit the students and may only hurt them through misinterpretation of the test results. Thus they are often reluctant to include their students.

However, exemption or exclusion of LEP students from testing is not the answer either. If they are not included in large-scale testing, then LEP students drop out of the accountability picture completely. If LEP students' progress is not followed, then institutions will not be held responsible for the students' performance. Particularly if there is interest in national measures of student performance to inform policy at the broader state or federal levels, LEP students cannot be exempted if the assessments are to be considered

as valid national indicators of student performance levels. If LEP students are included in large-scale assessments, however, there are further issues of how to ensure that assessment outcomes are appropriately reported and interpreted, and that they are used to ultimately benefit LEP students.

Part of this effort requires a clearer understanding on the part of all--those in the classrooms and schools, as well as those in district, state, and federal offices--regarding the purposes of specific assessments and the appropriate forms of assessment for different purposes. These concerns are addressed in the next section.

III. THE PURPOSES OF ASSESSMENT

Underlying the assessment issues presented in the previous section are certain assumptions about the purposes of assessment. It is broadly acknowledged that there are multiple purposes for LEP student assessment, but what is less likely to be acknowledged is that the types of assessment used for one purpose may not be useful or applicable for another purpose. In considering the value of assessment, in determining the appropriate approach to be used in assessment, and in designing its implementation, the purpose or purposes of a specific assessment should be of central importance. If the purpose(s) of a specific assessment are not kept clearly in mind, it is quite possible that the assessment will not provide the type of information needed.

There have been a number of attempts to categorize the purposes of LEP student assessment. In general those purposes can be categorized within two broad categories: (1) assessment focused on individuals; and (2) assessment focused on groups. In order to further clarify the role of purpose in this discussion of assessment of LEP students, nine major purposes of assessment within those two categories are defined. The purposes are as follows:

Individual

1. To initially determine whether a student is LEP -- **Identification**.
2. To assign a student to appropriate services -- **Placement**.
3. To determine a student's language-specific skills and weaknesses -- **Language Assessment**.
4. To regularly determine a student's academic progress, skills, and weaknesses relevant to the content curriculum -- **Academic Assessment**.
5. To review a student's service placement -- **Placement Review**.
6. To review a student's LEP status -- **Exit**.

Group

1. To assess class progress relevant to the curriculum -- **Instructional Assessment**.
2. To assess the effectiveness of particular instructional interventions or programs -- **Program Evaluation**.
3. To assess student achievement levels within specific schools or districts -- **Accountability**.

There are diverse audiences for information from assessment designed for these various purposes. Assessments for the purposes of identification, placement, placement review, and exit primarily serve school and district administrators. Assessments for the purposes of language assessment, academic assessment, and instructional assessment primarily serve teachers. Assessments for the purposes of program evaluation and accountability primarily serve education policy-makers at local, state, and federal levels.

Many analysts would suggest that **Identification** and **Exit**, and **Placement** and **Placement Review** have much in common, and thus could be combined in this categorization. Research evidence (e.g., Hopstock et al., 1993) suggests that different methods and procedures are used to address these purposes, however, and thus they are included separately in this presentation.

The importance of recognizing these distinct purposes of LEP student assessment becomes obvious when the characteristics of different assessments are described. Listed below are some characteristics or components of LEP student assessment which can vary across specific individual assessments, depending on the purposes.

Content Areas Tested

1. Testing of English language proficiency.
2. Testing of native language proficiency.
3. Testing of achievement in academic subjects

Nature of English Language Proficiency Testing

1. Testing of all modalities of proficiency (listening, speaking, reading, and writing).
2. Testing of academic language proficiency vs. basic interpersonal language skills.
3. Assessment of specific language skills and weaknesses.

Nature of Native Language Proficiency Testing

1. Testing of all modalities of proficiency (listening, speaking, reading, and writing).
2. Testing of academic language proficiency vs. basic interpersonal language skills.
3. Assessment of specific language skills and weaknesses.

Nature of Tests of Academic Achievement

1. Relevance of testing to actual school curricula.
2. Standardization of testing content across students/schools/districts.
3. Testing of higher level thinking skills vs. isolated facts.
4. Testing of knowledge and skills which students can apply to their lives.
5. Testing requiring student products (performance tasks, portfolios) vs. simple, stylized responses (e.g., multiple choice tests).
6. Testing of academic achievement in English.
7. Testing of academic achievement in the native language.

Test Characteristics

1. Documented reliability and validity of measurement.
2. Alternate forms/versions of test for multiple testing.
3. Limited testing/training time required.
4. Quick feedback on testing results to guide instruction.

In Table 1, the importance of these various characteristics/components for each of nine purposes of LEP student assessment (individual and group) is rated. The ratings are made based on the case of a bilingual education program in which LEP students are receiving instruction through significant use of their native

language. In the table, each characteristic of assessment is rated as very important (V), moderately important (M), or not very important (N) for each of the assessment purposes. The ratings given in the table represent one possible consensus (based on the authors' discussion); the specific ratings assigned are certainly arguable. The point of the table, however, is to illustrate that the importance of characteristics varies across assessment purposes. This is clearly seen by reading down one column for any one rating purpose and comparing the ratings with those assigned under another column, i.e., for a different purpose. It is clear that decisions regarding assessment can be very different when the purposes of the assessment vary.

There are at least three conclusions which should be drawn from the table. First, researchers and policy-makers who offer suggestions concerning LEP student assessment should specify the type of assessment to which they are referring. Suggestions concerning placement testing may have limited relevance for program evaluation or accountability purposes. Second, individuals designing LEP student assessments should be very conscious of the purposes of those assessments, and design assessments which are relevant to those purposes. For example, if quick feedback is required for the purpose of instructional assessment, a testing program without cumbersome scoring procedures or requiring specially trained raters should be used. Third, if a particular assessment is to serve more than one purpose, the tradeoffs among those purposes should be clearly recognized, and a consciously chosen compromise should be made. The optimal assessments for academic assessment are **not** the same as the optimal assessments for accountability, so if both purposes must be addressed by the same assessment, the tradeoffs should be carefully examined. Thus, for example, assessments relating to national standards may be intended to be used both for accountability and academic assessment purposes. In the absence of a mandated national curriculum, it is unlikely that these two purposes can both be effectively achieved with the same assessment.

A second and highly related issue which must be considered in examining any assessment system for LEP students is **what** it is appropriate to test. The content of a specific assessment may have varying relevance or appropriateness for students based on: (1) their language skills in English and the native language; (2) the instructional program which they are receiving (both in terms of content and the language(s) of instruction); and (3) the content and language used in the assessment. An assessment which may be appropriate for one student may be highly inappropriate for another student in the same school. Any assessment system must recognize this complexity, and make reasonable compromises between uniformity of assessment and appropriateness of methods.

These principles or cautions should be kept in mind as basic underpinnings of the discussion of assessment instruments and practices that is presented in the remainder of this paper. Indeed, the importance of considering the purpose of assessment is one that will continue to be seen in the discussions to follow.

The following section examines the nature and purpose of assessment as identified within effective programs, specifically, funded exemplary programs for LEP students. Then, in the next two sections, the most frequently used standardized achievement tests and language proficiency tests are examined and compared.

TABLE 1

The Importance of Various Assessment Characteristics for Different Assessment Purposes:

Programs with Primarily Native Language Instruction

(V = Very important, M = Moderately important, N = Not very important)

Assessment Characteristic	Individual						Group		
	Identification	Placement	Language Assessment	Academic Assessment	Placement Review	Exit	Instructional Assessment	Program Evaluation	Accountability
Content Area Tested									
Testing of English language proficiency	V	V	V	M	V	V	M	M	M
Testing of native language proficiency	M	V	V	M	V	M	M	M	M
Testing of achievement in academic subjects	M	V	M	V	V	V	V	V	V
Nature of English Language Proficiency Testing									
Testing of all modalities of proficiency (listening, speaking, reading and writing.)	M	M	V	M	M	M	M	M	M
Testing of academic language proficiency vs. basic interpersonal language skills.	V	V	V	M	V	V	M	M	M
Assessment of specific language skills and weaknesses.	M	M	V	M	M	M	M	M	N
Nature of Native Language Proficiency Testing									
Testing of all modalities of proficiency (listening, speaking, reading, and writing)	M	V	V	M	V	M	M	M	N
Testing of academic language proficiency vs. basic interpersonal language skills	M	V	V	M	V	M	M	M	M
Assessment of specific language skills and weaknesses.	M	V	V	M	V	M	M	M	N
Nature of Tests of Academic Achievement									
Relevance of testing to actual school curricula	M	V	N	V	V	M	V	M	M
Standardization of testing content across students/schools/districts	M	N	M	N	N	M	M	M	V
Testing of higher level thinking skills vs. isolated facts	N	M	N	M	M	N	M	M	M
Testing of knowledge and skills which students can apply to their lives	N	N	N	M	N	N	M	M	M

Testing requiring student products (performance tasks, portfolios) vs. simple, stylized responses (e.g., multiple choice tests)	M	M	M	M	M	M	M	M	M
Testing of academic achievement in English	M	M	M	V	V	V	V	V	V
Testing of academic achievement in the native language	M	M	M	V	V	M	V	V	M
Test Characteristics									
Documented reliability and validity of measurement	M	M	M	M	M	M	M	V	V
Alternate forms/versions of test for multiple testing	N	N	M	V	N	N	V	V	V
Limited testing/training time required	M	M	M	V	M	M	V	M	M
Quick feedback on testing results to guide instruction	N	N	M	V	N	N	V	N	N

IV. ASSESSMENT PRACTICES ASSOCIATED WITH SUCCESSFUL PROGRAMS

This chapter addresses the issue of whether there are particular assessment practices which were associated with successful instructional programs for LEP students. A review of the literature did not find any research or research summaries which were specifically devoted to this question. A number of sources (e.g., Council of Chief State School Officers, 1992) suggested guidelines for LEP assessment, but none of them empirically related those guidelines to effective programs.

One research study which examined programs that had been nominated as best implementers of specific program models in California (Gandara & Merino, 1993) did report on the availability of test data on students in those programs. The general finding of that study was that longitudinal test data for the purpose of program evaluation were frequently not available for analysis. The reasons for the lack of such data included mobility of students, changes in tests used, and absence during testing periods. This study, however, provided only very limited information on the overall assessment approach of the programs studied.

In order to gain further insight into this question, the applications from nine first year Title VII Academic Excellence projects funded in 1993 were reviewed. These projects were selected for funding based both on the educational significance and demonstrated effectiveness of the program and on the strength of the dissemination plan. In order to be funded, the program needed to be recognized as exemplary either by the U.S. Department of Education or by the relevant state education agency.

The review indicated that for the purpose of program evaluation, all of the projects had collected detailed test data to document their claims of effectiveness. In most cases this involved pre- and post-scores from the same students, and comparisons of those scores with national norms. In addition, all projects required adopting sites to provide similar test data in order to document effectiveness at non-model sites. The presence of these assessment elements was a requirement for participation in the Academic Excellence program, and thus were not necessarily considered to be component parts of the project "model."

Only one of the projects which were reviewed identified LEP student assessment as a distinct "key" component of the project which was to be disseminated. A project designed to identify and serve gifted and talented LEP students had student identification as a key component, and included new and adapted measures to be used for identification. In addition, a few projects included assessment instruments as part of their delivery system for instruction. For example, a project for kindergarten students included a "daily observation card" for each student for the purpose of recording progress toward instructional objectives. A project to improve writing instruction for LEP students in grades 4-12 included a newly developed computerized inventory to assess the development of writing skills.

What the review suggests is that LEP student assessment is typically thought of as an issue which is separate from that of effective instruction. It is likely that all of the projects had systems for identification, placement, language assessment, etc. of their LEP students, but the only assessment approaches which were described in detail were those which were required for purpose of program evaluation. If effective instruction and assessment are typically thought of as separate issues, this may suggest that additional attention should be paid to using assessment for instructional improvement purposes.

The review also highlights the importance of recognizing the different purposes of LEP student assessment, and designing an assessment system which meets those various purposes. The Academic Excellence program has specifically emphasized the program evaluation purpose, but other purposes might also be the focus of attention.

V. A REVIEW OF ENGLISH LANGUAGE PROFICIENCY TESTS

Administration of a language proficiency test in English is the most common method used to determine whether a student is limited English proficient (Hopstock, Bucaro, Fleischman, Zehler, and Eu, 1993). Eighty-three percent of school districts with LEP students were found to use this method, either alone or in combination with other techniques. Similarly, English proficiency tests are used by 64 percent of school districts for assigning LEP students to specific instructional services in schools, and by 74 percent of school districts for reclassifying students from LEP status.

The English proficiency tests most frequently used to identify, assign, and reclassify LEP students are the Language Assessment Scales (LAS), the Idea Proficiency Test (IPT), the Maculaitis Assessment Program (MAC), the Bilingual Syntax Measure (BSM), the Peabody Picture Vocabulary Test (PPVT), and the Language Assessment Battery (LAB) (Hopstock, et al., 1993). Locally developed tests are also frequently used. Often these tests are used in conjunction with information from other sources in making a decision about a student; however, especially in initial identification and placement, the English language proficiency test is a key factor.

These six most frequently used English language proficiency tests therefore were reviewed for the purposes of this report. These tests were compared for the theoretical bases underlying their construction, and for the

content and technical qualities of the tests.

A. Examination and Comparison of the Language Proficiency Tests

For the purposes of this report, we compared the test content and the nature of the items used for the six tests. In making these comparisons, where different levels/versions of a test were available, the version or level most appropriate for a third grade student was selected for this review. The specific tests reviewed were:

- Idea Proficiency Test-I
- Language Assessment Scales 1C (Oral)
- Language Assessment Battery IIA
- Bilingual Syntax Measure II
- Maculaitis Assessment Program, Level 2-3
- Peabody Picture Vocabulary Test-Revised.

1.Examination and comparison of item content and structure.

Table 2 presents a comparison in matrix form of the items presented in each test, including the specific content of the item and the type of skills assessed. Each item on all six tests has been categorized according to the nature of the skills required to complete the item. These categorizations of the different skills tested are based on our overall examination of the test items; they do not necessarily reflect the ordering or categorization presented in the tests.

The five categories presented in the table are: receptive skills (e.g., comprehension, discriminating sounds); oral production skills (e.g., producing words, sentences, phonemes either on the basis of nonlinguistic prompts or as repetition of a given prompt); combined receptive and production skills (e.g., items that require a student to understand a question or prompt and to produce an oral response); receptive and reading (e.g., the student listens to the item but reads and marks the correct answer); and reading only.

Some of the tests include a large number of items using a very limited number of elicitation techniques, while others consist of a few items each for a variety of techniques. The content comparison of the six language proficiency tests showed that the tests differ considerably in types of tasks and specific item content. Even where two tests appear to require the same type of response and similar item content, the scoring criteria may focus on totally separate aspects of the response. As a result, the items are actually assessing totally different skills.

As can be seen in Table 2, with the exception of the PPVT-R, the tests elicit oral production, for part or all of the responses. Many of the items require oral production in the form of naming. In the IPT I (10 percent of the items) and the LAS 1C (15 percent of the items) the student is asked to name a referent, activity or other aspect of a picture presented in the test. The MAC test has total of 40 items eliciting the name of a color, shape, or number. None of the other tests focuses specifically on these skills (although the PPVT includes items that refer to attributes).

Six percent of items on the IPT-I are scored based on word order while 35 percent of the items on the LAS 1C require correct repetition of phonemes. Both the IPT and the LAS include items on discrimination among minimal pairs. On the LAS, these are 35 percent of the total items, while on the IPT they are 11 percent. One other common elicitation technique used by the tests (e.g., IPT, LAB, and MAC) is open-ended responses to questions. Each question that is presented refers to a different topic; they do not utilize a

consistent discourse theme or context across a series of questions. The IPT-I contains the greatest number of this type of item format.

Items within the mixed category of "oral receptive and reading productive" are found in the LAB only. The LAB listening subtest has students respond to an oral item followed by the presentation of a set of oral responses. However, to respond, the students read responses and mark the appropriate one in the test booklet. Finally, of the six tests examined, only the LAB II and the MAC 2-3 included items categorized as reading skill items only.

TABLE 2

Comparisons of Six Most Frequently Used Language Proficiency Tests

A. Productive Skills

Tasks/Skills	IPT - I		LAS IC		LAB IIA		BSM II		MAC 2-3		PPVT-R	
	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Picture used as prompt to elicit:												
name of object, noun	4 3 3 2	C D E F	10	Part 1: Vocabulary (Name That Picture)								
name of color, shape, number									15	Part II: Vocabulary Knowledge		
irregular plural noun	1	C										
present progressive	1	C	10	Part 1: Vocabulary (Action Words)								
simple present tense	1	D							5	Part I: Oral Expression		
identify upper and lower case letters of alphabet												

* Blank cells indicate tasks/skills not included in the particular test listed.

Table 2

B. Receptive Skills

Tasks/Skills	IPT - I		LAS IC		LAB IIA		BSM II		MAC 2-3		PPVT-R	
	Items	Sub-	Items	Sub-Test	Items	Sub-	Items	Sub-	Items	Sub-Test	Items	Sub-

		Test			Test		Test			Test
Student points to picture to indicate:										
noun										X
adjectives										X
subject pronoun	1	C								
comparatives	1	D								
prepositions of place										
inflected verbs										X
possessive pronoun	1	D								
superlatives	1	E								
Repeat simple sentence with:										
correct word order	2	C								
	2	D								
correct phonemes (in a word, phrase or sentence)			35	Part 5: Phonemes						
Discriminate phonetic minimal pairs	4	E	35	Part 4: Minimal Sound Pairs						
	4	F								
Follow verbal command with non-verbal response	2	C								
	1	D								
Student identifies 1 picture described by examiner out of 3-4					16	Test 1: Listening Part I				
Word identification: student matches orally presented words								5	Part III: Listening Comprehension	
Student counts number of words heard								6	Part III: Listening Comprehension	
Sentence comprehension: student selects sentence closest in meaning to sentence heard.								5	Part III: Listening Comprehension	

Table 2

C. Receptive and Productive Skills

Tasks/Skills	IPT - I		LAS IC		LAB IIA		BSM II		MAC 2-3		PPVT-R	
	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Describe attributes of an object	3	E										
Answer open ended												

questions:												
without context	2	C			4	Test 4: Speaking			5 (Q/A read aloud by examiner)	Part III: Listening Comprehension		
	3	E										
	7	F										
within context	4	C										
	2	D										
Invent a story based on open-ended question	2	D										
Re-tell a story	4	F	X	Part 3: Story Retell								
Predicting using future tense	2	F						5				
Give opposite of word read aloud	2	E										
Student gives specific information/question as requested by examiner	2	D										
	1	E										
Use picture to:												
describe using past tense	1	E										
comparative	1	D										
predict using future tense	1	D										
Answer yes/no questions with context			10	Part 2: Listening Comprehension								
Respond to prompt and picture by indicating:												
location					2	Test 4: Speaking						
possessive					2	Test 4: Speaking						
present continuous					2	Test 4: Speaking	2					
past tense					1	Test 4: Speaking	7					
future					2	Test 4: Speaking						
has							1					
copula (sing. and pl.)							3					
plural							2					
article							1					
word order (embedded and							3					

simple question)												
reciprocal construction						1						
conditional auxiliary						2						
Answer WH - questions without context								5		Part I: Oral Expression		

Table 2

D. Receptive and Reading Skills

Tasks/Skills	IPT - I		LAS IC		LAB IIA		BSM II		MAC 2-3		PPVT-R	
	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Student selects correct written and oral response to oral WH-questions					14	Test 1 - Listening Part II						

Table 2

E. Reading Skills

Tasks/Skills	IPT - I		LAS IC		LAB IIA		BSM II		MAC 2-3		PPVT-R	
	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Cloze test in simple story:												
adjectives					9	Test 2: Reading						
nouns					14	Test 2: Reading						
verbs					13	Test 2: Reading						
Sentence completion with cloze					17	Test 3: Writing			10	Part II: Vocabulary Knowledge		
Answer question based on a preceding statement					2	Test 3: Writing			15	Part V: Reading Comprehension		
Identify a complete sentence					1	Test 3: Writing						
Alphabetize groups of words									6	Part IV: Word Recognition		
Identify single letters as vowels or consonants									6	Part IV: Word Recognition		
Use a labeled picture to determine whether named object in picture has a long or short vowel												

show singular or plural									6	Part IV: Word Recognition		
Add consonant to sequence of letters to form a word									6	Part IV: Word Recognition		
Select the word (out of three) with a silent letter									8	Part IV: Word Recognition		

TABLE 3**Six Most Frequently Used Language Proficiency Tests: Comparison of Administrative Procedures**

TEST	Grade level ranges?	Group or individual?	Time limits?	Taped stimuli?	Taped response?	Responses determine test length?	What's recorded by examiner?	Student completes answer sheet?
IPT-I	K-6	Individual	No	No	No	Yes	Correct or incorrect	No
LAS 1C	1-6	Individual	No	Yes	Optional for story retell	Yes	Incorrect only	No
LAB IIA	3-5	Group and Individual*, or Individual	Yes	No	No	No	---	Yes
BSM	K-12	Individual	No	No	No	No	Full student response	No
MAC 2-3	2-3	Group and Individual**	Yes	Optional	Optional for Part I-B	No	Proficiency rating for Part I-B only	Yes
PPVT-R	Ages 2.5-40	Individual	No	No	No	Yes	Correct or incorrect	No

*Requires an individual administration of one subtest (Test 4: Speaking).

**One subtest must be administered individually (Part I-B: Oral Expression).

2. Test Administration Procedures

Table 3 presents an additional example of the differences among the tests in their approach to assessing language proficiency. In this table, some key features of the administrative procedures of the tests are compared. The comparison shows that the tests differ in the grade level ranges for which they were designed. They also differ in whether there are specific time limits, in use of audiotaped information (either as prompts to elicit responses or as a record of the student's responses), and whether the test length is fixed versus determined by the student's responses. In addition, the tests vary in the procedures for recording responses and in the person doing the recording (examiner or student).

Differences among the tests in administration procedures demonstrate that it is equally important to note the format and procedures involved in presenting the test to a student as to examine the content of the items. For example, if it is optional to use the taped stimuli, then persons who administer the test within a district should decide to select a consistent option. Similarly, consistent decisions should be made regarding taping of students' responses.

Other aspects related to the administration of the tests include how responses are recorded. For example, in the LAS only incorrect responses are noted; this leaves open the possibility that the student will note the pattern and become intimidated by the examiner's markings. In the BSM, the examiner is to write down the student's full response. Beyond concern for how naturalistic a question and answer sequence can be under these circumstances, there is a concern for how accurately the examiner can record these responses. On the other hand, if students are to record their responses using answer sheets (as on the LAB, MAC), then additional test-taking skills are brought into play for the students in the assessment.

In some cases, the oral instructions given to the student prior to the actual item are more complex in language than the actual test items. This gives rise to the possibility that a student may misunderstand the instructions (or become flustered by not understanding them) and therefore miss the far simpler test item.

B.Theoretical Bases of the Tests

The language proficiency tests were each developed out of a current prevailing view of language acquisition processes. There is only a limited amount of overlap in the way in which language proficiency is defined and in the underlying approach to assessment of language proficiency or acquisition. Thus, each of the six represents a different conceptualization of the best means of identifying the level of proficiency of a LEP student.

In terms of current theoretical views of language proficiency and language acquisition, the Peabody Picture Vocabulary Test (PPVT-R) presents the most limited approach to assessment of language proficiency, especially oral proficiency. This test was originally developed much earlier than the other tests and prior to much of the basic research on first and second language acquisition processes. It reflects the viewpoint that acquisition is measured in terms of amount of vocabulary. (In fact, the earliest studies of child language acquisition involved word counts). In contrast to the approach taken in the PPVT-R, current views of language proficiency emphasize that language is an integrative skill, involving the ability to communicate effectively as a listener and speaker within actual language use situations.

The other five frequently used language proficiency tests reflect more recent views of language proficiency. In these, too, some of the historical development in the fields of first and second language acquisition research can be seen. For example, the Bilingual Syntax Measure was developed primarily on the basis of findings from first language acquisition research. In this research, sequences of acquisition of grammatical morphemes were identified (e.g., Brown, 1973) as occurring consistently across a number of children learning English as their first language. The common sequences observed across learners were interpreted as due to the operation of universal principles in language acquisition. Thus, the argument was made that the same universal principles could be applied to the second language learning of English by speakers of other languages (especially children). The Bilingual Syntax Measure was developed on the basis of this hypothesis that there is a consistent acquisition sequence of grammatical elements that can be used to assess level of acquisition. The test is designed to determine the level of proficiency based on the acquisition of specific syntactic forms for which a hierarchy of acquisition has been determined. The focus of the test, therefore, in both the item content and in the scoring guidelines is on syntactic competence. It differs from current views of language acquisition in this focus on syntax alone.

The technical description of the IPT-I also refers to first language acquisition. In the IPT, the sequence defined is a broader, developmental progression from babbling to echolalia, to telegraphic speech, to syntax. However, first language learners of any language generally move beyond two-word or three-word speech (the "telegraphic" stage referred to) at about age 2 or 3. The IPT-I assesses children from grades K-8. Thus, the test covers the period of "syntax" development; there is no specific definition of the approach to

assessment of language proficiency provided. Beyond the developmental outline of early language development, the IPT refers to the interaction of the child with the environment as a source of further language acquisition.

The MAC materials refer to a notional-functional approach to language proficiency; that is, the test is designed to emphasize "the vocabulary and structures needed by the language learner to respond appropriately in specific situations" (Examiner's Manual, p.3). Thus this implies an emphasis on contextualized use of language, within specific types of situations.

The LAB defines language proficiency in terms of communicative competence, i.e., as "the ability to receive and convey information through the use of oral and written language" (Abbott, 1985). Language proficiency is viewed as comprised of an underlying single unitary factor as well as some specific components that are unique to certain aspects of proficiency. For this reason, the approach to development of the LAB included assessment of separate components of listening, reading, writing, and speaking. The LAB is described as designed to assess language as an "integrative" skill (Oller, 1979) and to assess proficiency, especially as related to the New York City curriculum.

The LAS obtains a measure of oral language proficiency skills that would be needed in order to function effectively within an all-English mainstream classroom. The theoretical source of the LAS includes Cummin's distinction between decontextualized, academic language use and social, everyday, contextualized language use. The objective is to focus more on defining proficiency in terms of ability to function in classroom instruction where academic language proficiency is needed.

The six language proficiency assessment measures thus represent distinct approaches to definition of language proficiency, reflecting different theoretical emphases prevalent at the time of their development.

C. Issues Related to Validity and Reliability

The language proficiency tests reviewed for this report were also examined in terms of their validity as measures of proficiency and reliability in terms of the consistency in scores obtained. As already noted in some of the discussion above, the tests vary in the extent to which they provide a clear definition of the construct being assessed. Even when a clear rationale and description of the nature of language proficiency is provided, it may not fulfill all of the criteria that a teacher or researcher would require for language proficiency assessment. For example, in the PPVT-R, the measure of language ability is vocabulary comprehension, assessed by means of a student identifying the picture that corresponds to the object, attribute, or action given by the examiner. The test therefore involves only word comprehension and recognition skills, and does not require any oral production from the student. For this reason, the validity of the test as a measure of language proficiency, such as would be needed to communicate and discuss in the classroom, is weak.

Similarly, with regard to the BSM, it might be argued that the focus on syntax alone weakens the test's validity as a measure of oral proficiency over a range of uses and contexts. Thus, even though the BSM offers a research basis for the sequences assumed in item construction and scoring, the nature of language proficiency tested does not reflect a broad range of skills, that would be related to communicative competence in classroom and other discourse situations. The MAC, LAS, LAB, and IPT do utilize broader ranges of language skills to assess language proficiency. The LAB and MAC include items that tap literacy skills as well as oral proficiency skills. (The LAS-Oral but not the LAS-Literacy was included in this review; the LAS-Literacy was not among those instruments identified as most frequently used for identification and placement). Even these assessment instruments, however, do not fully assess a student's

ability to communicate meaningfully within actual discourse contexts, since they include items that predominantly test individual discrete language skills.

For several of the tests included in this review, past reviewers of the language proficiency tests have noted concerns with reliability and/or validity of the tests, the adequacy of the scoring directions, the limited populations on which test norms are based, and the availability of the conditions needed for administration of the measures. Reviews of the tests over the past several years have frequently noted such concerns regarding several of the tests and have recommended caution in their use in making educational decisions (e.g., BSM, LAS (and three other tests): Gillmore and Dickerson, 1979; BSM: Rosansky, 1979, Cziko, 1987; Hayes-Brown, 1987; Oller, 1976; IPT: McCollum, 1983, 1987; Rivera and Zehler, 1987; LAS: Haber, 1985). This caution is supported by other research showing that the tests do not agree in how they identify NEP/LEP/FEP (e.g., Ulibarri, Spencer, and Rivas, 1981; Gillmore and Dickerson, 1979). In addition, those who use the tests should be aware of the groups used to develop test norms. For example, the LAB is based on New York City students only and the norms are therefore limited in generalizability to other sites.

Considering the problems noted by several reviewers of the tests, the question that arises is why there has not been a greater debate about the nature and use of language proficiency tests (paralleling the concern regarding achievement tests described later in this paper). One explanation may be that the language proficiency tests generally have not been used for large-scale accountability purposes. The tests have been used to support decision-making for the purposes of identification, placement, and reclassification, typically in conjunction with other sources of information. Of the tests that are currently available, the Council of Chief State School Officers (CCSSO, 1992) have identified the LAS, LAB, and MAC as most closely meeting requirements stated in their recommendations on assessment of LEP students for these purposes (although they do not recommend use of these in isolation).

For reclassification purposes in the future, however, it will be increasingly important to assess LEP students' ability to communicate and participate meaningfully in the discourse-rich, hands-on classroom interactions defined by current instructional reform efforts. Assessment of language proficiency for these purposes will thus require the development of contextualized or authentic language tasks, or guidelines for more direct measures of communicative ability. The development of such assessments would more fully reflect current definitions of language proficiency.

VI. A REVIEW OF MOST FREQUENTLY USED STANDARDIZED ACHIEVEMENT TESTS

According to Hopstock, Bucaro, Fleischman, Zehler, and Eu (1993), achievement tests in English are frequently used by school districts and schools to help identify LEP students, assign them to school services, and reclassify them from LEP status. Specifically, 52% of school districts and schools across the country use achievement tests in English to help identify LEP students. Approximately 40% of districts and schools use achievement tests to help assign LEP students to specific instructional services within a school, and over 70% of districts and schools use achievement tests to help reclassify students from LEP status. In general, the more LEP students a district has, the more likely it is that achievement tests in English will be used to make these important decisions. Hopstock et al. (1993) report that the achievement test batteries most frequently used by school districts for these purposes are the Iowa Test of Basic Skills (ITBS), Stanford Achievement Test (SAT), Comprehensive Test of Basic Skills (CTBS), California Achievement Test (CAT), and the Metropolitan Achievement Test (MAT).

All five achievement tests were designed to measure the learning outcomes of school curricula in the areas

of reading, language arts, math, science, and social studies. The publishers of the tests state that this objective was accomplished by reviews of major textbook series; samples of curriculum materials and syllabi and the National Council of Teachers of Mathematics (NCTM) Standards; and inputs from nationally recognized leaders in education concerning currents and emerging curriculum trends across the country. All five tests are designed to be administered in group settings, usually by classroom teachers, using standardized procedures. The exact instructions to be read to the students are printed in administration manuals, and students respond to multiple choice-style items directly in the test booklet or on separate answer sheets.

The technical manuals do not discuss the appropriateness or inappropriateness of the tests for LEP students. In general, LEP students do not appear to be included in the groups on which the tests were normed. The CAT and CTBS, however, did report the percentage of students in the norm groups of schools who came from "homes in which a language other than English is spoken most of the time." The percentages of such students were 8.2% for the CAT and 10.0% for the CTBS. The percentage of these students who were LEP and were included in the standardization samples were not discussed. Small percentages of LEP students were included in the ITBS standardization sample, ranging from a high of 1.5% in Grade 6 to a low of 0.3% in Grade 12. Students enrolled in bilingual or ESL programs ranged from 1.7% of the ITBS national standardization sample in Grade 3 to 0.5% in Grades 9, 10, and 12. No mention of LEP students being included in standardization samples was made in the SAT and MAT technical manuals.

The five achievement test batteries were reviewed in order to examine and compare the skills tested on each. Since each battery contains different sub-tests and levels for grades K-12, it was decided to confine the reviews to the level appropriate for the second half of grade 3. The editions, forms, and levels of the test batteries which were included in the reviews are as follows:

SAT, Eighth Edition, 1992, Form L, Primary 3;
 MAT, Seventh Edition, 1993, Form S, Elementary 1;
 CAT, Fifth Edition, 1992, Form A, Level 14;
 CTBS, Fourth Edition, 1989, Form A, Level 14; and
 ITBS, 1993, Form K, Level 9.

Table 4 shows the number of items and names of the sub-test in which the items are found within each of the five test batteries for specific tasks/skills in reading, math, language, study skills, listening, science and social science. A comparison of the similarities and differences in the tasks/skills found across the five batteries may be found below.

Reading

- All test batteries include significant numbers of reading comprehension items.
- The SAT and ITBS include more items calling for the identification of synonyms than do the other test batteries. On the other hand, the MAT, CAT, and CTBS include items calling for identification of opposites, while the SAT and the ITBS do not.
- The SAT includes items calling for "word study" skills such as dividing words into syllables and identifying letter-sound correspondence. None of the other test batteries include such items.
- The SAT includes items concerning words with multiple meanings. The other tests do not include such items.

- The MAT, CAT, and CTBS include items calling for choosing words to properly complete a sentence. The SAT and ITBS do not include such items.
- The CAT and CTBS include items which require identifying the meaning of prefixes and suffixes, and the identification of modern English words from non-English words; the others do not.

Math

- All test batteries include significant numbers of items calling for the addition, subtraction, multiplication and division of whole numbers.
- Only the CAT and CTBS have items involving addition and subtraction of fractions and decimals. However, all test batteries include items concerning the understanding of the meaning of fractions.
- All test batteries include items involving numerical word problems and the interpretation of tables and graphs.
- All test batteries include items involving principles of geometry.

Language

- All batteries contain items concerning the spelling of words.
- All batteries contain items concerning correct capitalization, punctuation, and usage.
- All batteries contain items concerning identification of complete and correct sentences. Except for the ITBS, the batteries contain items calling for correctly joining two sentences and/or rewriting sentences correctly.
- The CAT, CTBS, and ITBS include items which call for an understanding of paragraphs and how to use sentences within paragraphs. The SAT and MAT do not include such items.

Study Skills

- All test batteries include items calling for the identification of appropriate reference materials, and identifying information from tables of contents, indices, dictionaries, and encyclopedias.
- The SAT and ITBS include items on alphabetic ordering of words. The MAT has only one such item; the CAT and CTBS do not have any.

Listening

- Only the SAT has items involving listening. These items include identifying synonyms from key words in sentences read aloud by the teacher, and answering questions about a paragraph read aloud by the teacher.

Science/Social Science

- All test batteries include items concerning science and social science facts and interpretation.
- All test batteries include items calling for interpretation of tables and graphs.
- All test batteries include items on map reading.

As these comparisons show, there is considerable overlap among the tests in the types of skills assessed. Some differences emerge, however, and have implications for selecting assessment instruments. Comparisons among tests are important for administrators and assessment specialists to carry out to determine the degree of match between the instruction received by students and the skills required by the tests.

TABLE 4**Comparisons of Five Most Frequently Used Achievement Tests****Reading Tasks/Skills Included Within Five Achievement Test Batteries***

(* Test levels appropriate for end of third grade)

Tasks/Skills	SAT		MAT		CAT		CTBS		ITBS	
	Items	Sub-test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Identify synonyms/definitions	28	Reading Vocabulary			6	Vocabulary	7	Vocabulary	26	Vocabulary
Identify synonyms using clues in sentence	6	Reading Vocabulary	12	Reading Vocabulary						
Identify opposites					5	Vocabulary	5	Vocabulary		
Identify opposites using clues in sentence			8	Reading Vocabulary						
Divide words into syllables	12	Word Study Skills								
Identify sounds of letters within words	36	Word Study Skills								
Identify meaning of prefix and suffix					5	Vocabulary	6	Vocabulary		
Identify modern word from non-English word and its meaning					3	Vocabulary	3	Vocabulary		
Chose word that completes sentence			10	Reading Vocabulary	21	Vocabulary	19	Vocabulary		
Identify sentence in which multiple meaning word has same meaning as in first sentence	6	Reading Vocabulary								

Answer questions about paragraphs	54	Reading Comprehension	67	Reading Comprehension/ Language	50	Comprehension	57	Comprehension/ Social Studies	36	Reading Comprehension
Literal comprehension										
Inferential comprehension										
Critical comprehension										

Math Tasks/Skills Included Within Five Achievement Test Batteries*

	SAT		MAT		CAT		CTBS		ITBS	
Tasks/Skills	Items	Sub-test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Understanding arithmetic operations	10	Concepts of Number	5	Math Concepts and Problem Solving	2	Math Concepts and Applications	1	Math Concepts and Applications	1	Math Concepts and Estimation
Naming and reading numbers	7	Concepts of Number	2	Math Concepts and Problem Solving	2	Math Concepts and Applications	1	Math Concepts and Applications	1	Math Concepts and Estimation
Place value	2	Concepts of Number			2	Math Concepts and Applications	1	Math Concepts and Applications		
Ordering of numbers	8	Concepts of Number	3	Math Concepts and Problem Solving	5	Math Concepts and Applications	5	Math Concepts and Applications	2	Math Concepts and Estimation
Estimating values (rounding)	3	Concepts of Number	3	Math Concepts and Problem Solving	2	Math Concepts and Applications	2	Math Concepts and Applications	13	Math Concepts and Estimation
Addition/subtraction of whole numbers	24	Math Computation	5	Math Procedures	12	Math Computation	12	Math Computation	26	Math Concepts and Estimation/ Math Computation
Multiplication/division of whole numbers	20	Math Computation	7	Math Procedures	17	Math Computation	17	Math Computation	11	Math Concepts and Estimation/ Math Computation
Identification of odd/even	1	Concepts of Number	2	Math Concepts and Problem Solving	1	Math Concepts and Applications	2	Math Concepts and Applications	1	Math Concepts and Estimation
Understanding fractions	3	Concepts of Number	2	Math Concepts and Problem Solving	2	Math Concepts and Applications	2	Math Concepts and Applications	1	Math Concepts and Estimation
Addition/subtraction of fractions					6	Math Computation	6	Math Computation		
Understanding decimals			1	Math Concepts and Problem Solving	1	Math Concepts and Applications				
Addition/subtraction of decimals					9	Math Computation	9	Math Computation		
Knowledge of			4	Math Concepts	2	Math Concepts	2	Math Concepts	2	Math Concepts

measures				and Problem Solving		and Applications		and Applications		and Estimation
Understanding probability			1	Math Concepts and Problem Solving	1	Math Concepts and Applications	1	Math Concepts and Applications		
Numerical word problems	21	Math Applications	18	Math Concepts and Problem Solving/Math Procedures	10	Math Concepts and Applications	17	Math Concepts and Applications	17	Math Concepts and Estimation/ Math Problem Solving and Data Interpretation
Interpreting tables/graphs	8	Math Applications	4	Math Concepts and Problem Solving	11	Math Concepts and Applications/ Study Skills	10	Math Concepts and Applications/ Study Skills	10	Math Problem Solving and Data Interpretation
Value of money	2	Math Applications			2	Math Concepts and Applications			1	Math Concepts and Estimation
Reading clocks	2	Math Applications	1	Math Concepts and Problem Solving	1	Math Concepts and Applications				
Principles of geometry	5	Math Applications	6	Math Concepts and Problem Solving	6	Math Concepts and Applications	7	Math Concepts and Applications	3	Math Concepts and Estimation
Spacial relations					3	Math Concepts and Applications	1	Math Concepts and Applications	1	Math Concepts and Estimation
Identify misspelled word									27	Spelling
Identify misspelled word in sentences	36	Spelling	9	Language	30	Spelling	4	Spelling		
Choose word that is spelled correctly							26	Spelling		
Identify correct/incorrect capitalization	5	Language Mechanics	2	Language	10	Language Mechanics			24	Capitalization
Identify correct/incorrect punctuation	8	Language Mechanics	3	Language	26	Language Mechanics	4	Language Mechanics	24	Punctuation
Identify correct capitalization and punctuation							32	Language Mechanics		
Identify correct abbreviation	1	Language Mechanics								
Identify correct/incorrect usage	4	Language Mechanics	6	Language	5	Language Expression	12	Language Expression	19	Usage/ Expression
Identify correct verb tense	4	Language Mechanics	5	Language	4	Language Expression			2	Usage/ Expression
Identify error as	8	Language								

capitalization, punctuation, usage		Mechanics								
Identify parts of speech					5	Language Expression	4	Language Expression		
Identify complete and correct sentence	20	Language Expression			13	Language Expression	13	Language Expression	6	Usage/ Expression
Rewrite sentence correctly			4	Language						
Correctly join two sentences	10	Language Expression	3	Language	8	Language Expression	7	Language Expression		
Choose topic sentence in paragraph					3	Language Expression	3	Language Expression	1	Usage/ Expression
Choose sentences that develop topic sentence					4	Language Expression	3	Language Expression	1	
Reorder sentences in paragraphs									1	Usage
Choose sentence that does <u>not</u> belong in paragraph					3	Language Expression	3	Language Expression	1	Usage/ Expression
Choose sentence that belongs in paragraph					3	Language Expression	3	Language Expression		
Determine alphabetical order of words	7	Study Skills	1	Language					6	Reference Materials
Identify information from title page/table of contents/index	8	Study Skills	4	Language	7	Study Skills	4	Study Skills	6	Reference Materials
Identify information from library catalog card					3	Study Skills	5	Study Skills		
Identify information from sales catalog/poster					4	Study Skills				
Identify appropriate reference book to use	3	Study Skills	4	Language	2	Study Skills	1	Study Skills	1	Reference Materials
Obtain information from dictionary/encyclopedia	9	Study Skills	1	Language	5	Study Skills	4	Study Skills	6	Reference Materials
Identify headings/information from outlines/word lists	3	Study Skills			2	Study Skills	5	Study Skills		
Identify key words from sentence for research notes							1	Study Skills		
Identify uses of reference materials									4	Reference Materials
Identify key words and									4	Reference

locate appropriate card catalog drawer										Materials
Identify appropriate section in library							1	Study Skills	1	Reference Materials
How to organize research paper					1	Study Skills				
Listening Tasks/Skills Included Within Five Test Batteries*										
	SAT		MAT		CAT		CTBS		ITBS	
Tasks/Skills	Items	Sub-test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test	Items	Sub-Test
Identify synonym from key word in sentence read by teacher	15	Listening								
Answer questions about paragraph read by teacher	30	Listening								
Science facts/inference	36	Science	24	Science	17	Science	33	Science	30	Science
Interpretation of data	14	Science	11	Science	23	Science	7	Science		
Social science facts/inference	41	Social Science	28	Social Studies	13	Social Studies	9	Social Studies	26	Social Studies
Map reading	7	Social Science	4	Social Studies	11	Study Skills/ Social Studies	19	Study Skills/ Social Studies	16	Social Studies/ Maps & Diagrams
Analysis of graphs/charts	2	Social Science	3	Social Studies	18	Social Studies	12	Social Studies	12	Maps & Diagrams

* Test levels appropriate for end of third grade

VII. ISSUES IN THE USE OF STANDARDIZED TESTS

Over approximately the past forty years, there has been an increasing use of standardized testing (especially standardized academic achievement tests), and an expansion in the purposes for which tests are administered (Haney and Madaus, 1989). As the quantity of testing has grown, however, there has also been an increasing dissatisfaction with several aspects of the tests and their uses (Geisinger, 1992; Gandara and Merino, 1993; Haney and Madaus, 1989; Mehrens, 1992; Shepard, 1989). Many of these concerns have been voiced regarding the use of the tests with the general population of students. However, the same issues apply to the use of the standardized tests with language minority and LEP students, and have perhaps even greater implications for these students.

A. Criticisms of Standardized Achievement Tests

Arguments against the use of standardized achievement tests have addressed effects of the tests on instruction, on selection and placement, and on the uses of test results for accountability purposes (Darling-Hammond, 1994; Haney and Madaus, 1989; Milk, 1993; Shepard, 1989; Worthen, 1993). They have also related to the appropriateness of the instruments for LEP students based on language and cultural considerations. The specific criticisms related to the use of standardized tests are presented below.

1. Tests focus on lower order skills and lead to narrowing of the curriculum.

Standardized tests are most often multiple-choice tests. While this format does not preclude items that can tap use of higher order skills, most often the types of skills that are required are lower-order skills such as recognition of a correct response (as opposed to higher order tasks, e.g., tasks that require application of knowledge to solve a problem). Since current research emphasizes the importance of promoting higher order thinking skills, the standardized multiple choice tests are criticized for their failure to assess these critical cognitive skills. The argument is made further that, given their focus on lower order skills, the tests provide a poor indication of a student's true level of understanding of the content being tested.

A closely related criticism of the use of standardized tests is based on the unintended outcomes of the tests. Particularly when there are high stakes attached to the results of tests, teachers tend to "teach to the test". That is, teachers begin to teach to the test in order to prepare students (Meisels, 1989; Milk, 1993; Shepard, 1989; Shohamy, 1993). As a consequence, what is taught to students takes on the same emphasis on lower-order, rote skills as is present in the test, and time for other instruction is lost to test preparation (Darling-Hammond, 1994).

2. Test results are used to make inappropriate placements and decisions concerning LEP students.

Tests are also used to determine students' eligibility for special programs, and to identify competence in skills required for promotion or graduation. These are also "high-stakes" testing purposes in that performance on the test determines future prospects for the student (Madaus, 1989, cited in Sosa, 1990). Assessment outcomes may limit the students' potential to achieve at high levels. For example, when students are placed within lower instructional tracks, the instructional content they receive, is much more likely to be focused on lower order cognitive skills, such as in drill and rote learning tasks. Unfortunately, poor and minority students are found to be selected for lower track placement more often than would be expected based on their percentages in the population (Oakes, 1990).

Misclassification of LEP students often occurs when those administering the assessment lack an understanding of the effect of second language acquisition patterns on assessment outcomes (Shiff-Myers, Djukic, McGovern-Lawler, and Perez, 1992; Ascher, 1990; Bermudez and Rakow, 1990; Ortiz and Wilkinson, 1990). Thus, LEP students are overrepresented in special education classes but underrepresented in gifted and talented classes (Darling-Hammond, 1994; Ortiz, 1992), due to the lack of understanding about the role of language.

In addition, for individual students, the outcomes of a high stakes test can have a significant impact on that individual's future. For example, where eligibility to graduate is determined by test scores (Airasian, 1987; Darling-Hammond, 1994), LEP students who enter at the secondary level often do not have the level of English proficiency required to pass the required minimum competency exams in English. Without a diploma, these students will face later difficulties in obtaining employment.

3. Test are inappropriately used for evaluation/accountability.

The uses to which test results are put are also a critical issue. As Airasian (1987) points out, tests have become important for decisions involving staffing, comparisons among schools and districts, funding decisions, and program evaluation. Program outcomes in terms of student scores on achievement tests have been used as data in evaluation of programs, and as accountability data in assessing teachers, schools, districts.

Despite the increased emphasis on testing over the past two decades, test data are often not consistently available for LEP students (Gandara and Merino, 1993). Often LEP students are either exempted from taking the test (LaCelle-Peterson and Rivera, 1994) or the LEP students take the test but their scores are not reported or included within any local averages. Many teachers and administrators feel strongly that their students should not be required to take the test out of a concern for the student, whose English proficiency may not be adequate. In addition, there is a further concern regarding the effect of LEP students' scores on the overall means on the test. Spencer (1994) has presented data showing at the state level how inclusion of LEP students can affect the overall scores reported; for example, with LEP students included, the State of California drops significantly in overall ranking on test scores for NAEP.

Also, there are also concerns that the test results for groups may be misinterpreted. Hakuta (1986) has reviewed past testing practices with language minority populations showing that misuse and misinterpretation has often occurred. It is in part this type of historical record that overshadows some of the current debates about LEP students in assessment and that has created a reluctance towards assessment on the part of teachers and other staff who work with LEP students.

When LEP students are included in testing in English, analysts do not always recognize that the outcomes reflect the operation of two factors: language proficiency and academic knowledge. Thus, LEP students' achievement is underestimated (LaCelle-Peterson and Rivera, 1994; Secada, 1994). Comparison of LEP students' scores against test norms is questionable, given that the norms generally are based on mainstream, monolingual or English-proficient students (LaCelle-Peterson and Rivera, 1994). This type of comparison is not an equivalent one. For LEP students, the assessment will test both language and content; for the monolingual student, the assessment will measure content knowledge, and the role of language proficiency will be minimal.

However, exclusion of LEP students from testing due to their low level of proficiency in English is also not appropriate if schools and districts are to be held accountable for the performance of all of their students (Canales, 1994). The issue is therefore how to meaningfully include LEP students within assessment for accountability so that their needs and performances are recognized while ensuring that there is no misuse or misinterpretation of the LEP students' scores.

4. The roles of language and culture in test performance are not recognized.

Research on language acquisition has demonstrated that second language acquisition is a long-term process (Collier, 1992) and that fluency with the language in social interaction does not necessarily indicate full acquisition of decontextualized, academic language use (Cummins, 1984).

Particularly when tests are used for the purpose of examining growth in student achievement, the question of student language proficiency presents complications. The LEP student who is gaining in content knowledge through instruction in English is also gaining in English language proficiency, a process which requires somewhere between 4-8 years for full proficiency to be achieved (Collier, 1992). How can increases in scores on a test of content area be interpreted: Do they represent gains in content area knowledge? Or, do they represent gains in English proficiency? Or, perhaps most accurately, what portion of an increase in score should be attributed to these two factors?

The role of culture is also important. In current descriptions of the learning process, learning is described as "constructive," that is, built or created by the learner on the basis of existing or prior knowledge (e.g., Anderson et al., 1985). This includes much of a learner's knowledge that is built upon day-to-day experience. When a lesson, reading text, or test item is based on objects, situations, or viewpoints that are

particular to only certain groups, then students outside those groups will be at a disadvantage. The students who do not share the experience assumed by the text will have difficulty in understanding and learning from the text, or in demonstrating their ability in skills/concepts through the use of a specific test item. Thus, cultural differences can affect the level of performance of language minority students on tests which have been developed primarily for mainstream students (Geisinger, 1992; Hamayan and Damico, 1991; Madaus, 1994; Navarette et al., 1990).

B. Concerns Regarding the Language of Administration in Standardized Testing

An important issue that arises in designing assessment for LEP students is whether to test in the student's native language or in English. In order to obtain a more fair test of a LEP student's level of achievement, one recommendation is to test the student in his/her native language. This recommendation, however, leads to other issues. For many language groups, the resources for testing in the students' native language are not available and use of English is the only possible option. The question then is whether assessment can be administered in the native language for some language groups but not others.

Second, LEP students may have oral skills in their native language, but these may be weak or in the process of attrition. This can occur, for example, when a student is in a point of transition from the native language to the second language, English. Even where oral native language skills are strong, the students may lack needed literacy skills in their native language. What is then the appropriate approach to assessment for these students?

Third, even when a student is proficient in the native language, the language of assessment must also be related to the language used for instruction. That is, is a test of mathematics given in Spanish to students whose instruction has been all in English any fairer to the student? If a student is tested in the native language and is taught in the native language with a transition to English, at what point is assessment also transitioned into English?

The answers to questions related to LEP students' inclusion in testing, whether in English or in their native language will not be simple. In addition, these answers will need to be formulated with a clear understanding of the purposes of the assessment and of the factors involved in LEP students' performance. Once questions about language of administration are answered in principle, further decision-making will need to address the practical questions regarding implementation.

If it is decided in principle that native language testing will be used where appropriate, then this decision sets in place a series of implementation issues. These include outlining the process and criteria for identifying those students for whom native language tests are the most appropriate (e.g., language dominance testing, review of individual students' instructional programs, etc.). In addition, how to schedule and implement the testing will need to be determined as well as how and when to transition a student to English language tests. Each decision should be made with the realization that at the school and classroom levels any one requirement may involve a significant commitment of resources. Thus implementation concerns should be a critical component of decision-making regarding assessment, especially regarding the issue of the language of administration of a particular test.

VIII. ALTERNATIVE ASSESSMENT

There has been increased interest in alternative means of assessment, especially to assess higher order cognitive skills and application of knowledge to new problem-solving situations. For LEP students in

particular, alternative assessment approaches have been viewed as very useful new directions to explore.

A. Definition of Alternative Assessment

Alternative assessment perhaps is most simply defined as being different from standardized, multiple choice or discrete point tests which test skills through decontextualized items and which generally focus on lower order skills. In contrast to standardized tests, alternative assessment approaches:

- tap use of higher order thinking skills;
- require reasoned application of acquired knowledge rather than rote repetition of facts or formulae;
- are defined within contexts that are meaningful to students, either as relevant to real-world tasks and/or as examples of ongoing work produced by the student.

Alternative assessment approaches generally involve more extended tasks, i.e., tasks which require more time to complete or that involve a series of multiple tasks (Baker, 1992). Alternative assessment can also include student-constructed products that demonstrate the student's knowledge or proficiency. Such demonstrations might be a specific performance or product (e.g., an oral speech, a play about an historical event) or, the product could be a longer term demonstration such as a portfolio that represents examples of student work over time (e.g., writing samples, examples of problem-solving in mathematics). Alternative assessments can be both structured approaches (e.g., checklists, rating scales) as well as unstructured approaches (e.g., writing samples, anecdotal notes) (Navarette, Wilde, Nelson, Martinez, and Hargett, 1990).

A distinction is generally made between two types of assessment: performance assessment and portfolio assessment (e.g., Valdez-Pierce and O'Malley, 1992). Performance assessment (e.g., oral interviews, story retelling, simulations, hands-on lab work) involves a student's completion of a task to demonstrate specific skills and competencies in relation to agreed upon standards (Valdez-Pierce and O'Malley, 1992; Feuer and Fulton, 1993). The student's performance is observed and rated according to specified criteria. Portfolio assessment is an ongoing, systematic and purposeful collection of student work that reflects progress toward specific goals. Portfolio assessment often includes student self-reflection and monitoring as key elements (Valdez-Pierce and O'Malley, 1992).

For purposes of guiding classroom instruction, teachers and students can make selections of tasks, form judgements about the work, and identify areas which need further effort. However, in large-scale uses of alternative assessment, the selection of tasks, judgements of performance, and interpretation would be carried out outside of the classroom, by trained raters.

B. Motivations for a Shift Toward Alternative Forms of Assessment

Alternative assessment has begun to be viewed as an important direction in assessment that is consistent with much of what is being emphasized in the research on learning processes and on effective instruction.

1. Restructuring of viewpoints on learning processes and assessment.

The objective of instructional reform efforts overall has been the development of more challenging and meaningful learning tasks for students that promote the development of higher order thinking skills. In research on learning processes, effective instructional models are further defined as those in which students are active participants in instruction, taking responsibility for their own learning, and teachers are facilitators

of this process (e.g., Brown et al., in press; Resnick, 1987; Special Issues Analysis Center, 1993; Warren and Rosebery, 1990). This model is derived from research in which learning is defined as the product of interaction of a learner with others who guide the development of skills and understanding, such as outlined in the research of Vygotsky (1978, "Zone of Proximal Development").

Within the area of instructional reform, recent work is also focused on the creation of more "authentic" instructional contexts (Newmann, F.M. and Wehlage, 1993) in which students develop new knowledge through application to problems that have meaning to them, i.e., problems that are similar to those they would encounter outside the classroom. Multiple-choice tests are viewed as generally failing to assess the higher order thinking skills emphasized by these areas of research, and thus there has been growing interest in alternative and more "authentic" assessment (Wiggins, 1989) tasks.

In the area of language proficiency, language use is described as an integrated process rather than a composite of discrete skills (Oller, 1992) and there is an emphasis on the highly contextualized nature of language learning and language use (McLaughlin, 1987). Good assessment of language proficiency therefore examines language use within more contextualized and meaningful tasks (Bachman, 1990), rather than examining discrete, decontextualized language skills (e.g., such as grammatical forms in isolation).

2. The use of alternative assessment to promote improvement in instruction

Alternative assessment is viewed by many as a means of promoting school change (Darling-Hammond, 1994), and promoting effective instruction. It is thus seen as a specific tool to be used in educational reform. The use of alternative assessment approaches is expected to promote the use of participatory or active instructional settings. Such assessments are expected to provide to both teachers and students information that is directly relevant to their ongoing work in the classroom, indicating level of skills and content knowledge to guide further learning efforts. For LEP students, the flexibility of alternative assessment formats and the potential for structuring these to adapt to the language in which the student is learning are seen as offering a more appropriate means of measuring progress toward academic goals. Thus, especially in terms of assessment used for the purpose of guiding instruction, alternative assessment offers important opportunities for LEP students in particular.

3. Alternative assessment to inform high-stakes decision-making.

Alternative assessment approaches are being examined as effective means of assessing higher order thinking skills such as will be incorporated within proposed national standards. Alternative assessment tasks are viewed as more appropriate for these purposes since they provide different formats, extended time for in-depth work, and student direction of their own efforts, in contrast to multiple-choice tests. However, as with current multiple-choice tests, issues of reliability, validity, and generalizability must be addressed. Similarly, issues of how to fairly include LEP students must be addressed.

C. Issues in the Development of Alternative Assessment Approaches

Alternative assessment is gaining much support. For example, the Office of Technology Assessment (1992) reported that as of 1991, thirty-six states were using writing assessments and nine other states were in the planning stages. Of the 36 states, 21 used at least one other type of performance assessment (e.g., constructed responses, portfolios, hands-on demonstrations). Most of these states were using the new tests in conjunction with multiple-choice tests. However, many of the same issues for which the standardized multiple-choice achievement tests are criticized can be equally problematic for alternative assessments (Baker, 1992; Mehrens, 1992). In addition, other concerns may arise due to the nature of the tests (For

example, development and maintenance costs may be a large issue; Mehrens, 1992).

1. What is required for reliable and stable judgements?

Alternative assessment tasks generally involve a demonstration or product which is to be observed and rated by a judge. For example, writing samples or answers to a science problem must be rated following stated criteria. Individual judges who rate the same performance or product should agree on their ratings, and ratings from assessment of a particular skill within one task should agree with ratings of the same skill from another task. Thus far, the results on reliability are mixed. The Vermont program of assessment has resulted in portfolio outcomes which show promise but which for now do not provide sufficient reliability ratings to support their use in large-scale accountability studies.

Reliability of performance assessment has been primarily reported for writing assessment, for which rater reliability is generally in the low .80s (Mehrens). Interim results of the State of Vermont's portfolio assessment program have shown some increases in reliability of ratings in the second year of implementation for portfolios in writing and in math; however, while reliability for mathematics was reasonably strong, particularly in grade 8, the reliability for writing portfolios was not satisfactory and indicated further attention to the definition of the writing program (Koretz et al., 1993).

2. Can validity of the tasks be demonstrated?

Alternative assessments have face validity (Mehrens, 1992). That is, they appear to measure the skills that educators want to measure. For example, in the area of written work (the most common form of performance assessment involves writing), alternative assessment offers direct observation of the skills to be measured: integrated use of language to communicate with the reader. However, further research is needed to examine the actual validity of the alternative assessment tasks for measuring the higher order skills they are intended to measure (Mehrens, 1992).

One important problem related to validity is the limited task sampling that is possible given the fact that each performance assessment task requires an extended time period. Generalizability from a limited set of specific tasks to a larger curricular domain is a question. Thus far, it appears that results of assessment of higher order cognitive skills are limited in their generalizability (Mehrens, 1992; Miller and Legg, 1993; Shavelson, Baxter, and Gao, 1993).

The time and resources required for development and administration of alternative assessment tasks limit the number of topics or content areas that can be assessed. This being so, there is a concern that ultimately the use of alternative assessments for purposes of accountability could narrow classroom instruction by limiting instruction to those skills and topics included in the assessment.

D. Issues in the Implementation of Alternative Assessment

Standardized, multiple-choice tests have an advantage in terms of cost and time compared to alternative assessments. Both in terms of development and implementation, alternative assessments are likely to require a very lengthy process to define, develop, administer, and interpret. Before use of alternative assessment can be put into place successfully, there are several issues to be addressed related to their implementation. These are outlined below:

1. Who administers and rates?

Alternative assessments used for instructional purposes ultimately rely on users who will be in the classroom--the teacher. As Rueda points out (1992), it is critical to consider the "end-users" of the assessment measures being developed. In particular, it is important to recognize that the alternative assessments being proposed represent a major departure from the assessments that have been used in the past. Most of the teachers who will be using these assessments have been trained within the older assessment framework. If alternative assessments rely on teachers, then their effective use will be dependent upon teachers' willingness to change. Real conceptual change is not easy to achieve and teachers need ongoing assistance in that process. Yet if teachers are not able to change their beliefs, few changes in practices will occur (Peterson, 1992). For alternative assessment used with LEP students, the persons who develop and implement the review, scoring, and interpretation of student outcomes will need to be informed regarding language minority issues and the role of language acquisition in performance.

Similarly, for alternative assessment used for large-scale accountability purposes, those who develop and implement the assessments will need to consider the issues involved for language minority and LEP students related to review, scoring, and interpretation of student products.

2. Is the assessment task format related to classroom tasks?

Students' lack of familiarity and comfort in working within a performance task setting will hinder their ability to complete the task, compared to students who have prior experience in working with these tasks. Shavelson, Baxter, and Pine (1991, in Ruiz-Primo, Baxter, and Shavelson, 1993) found an effect of student familiarity with hands-on science tasks in their research on performance assessment in science. The researchers found a consistently higher mean performance score on hands-on science tasks for students who were experienced in these types of tasks compared to students who were not experienced. This linkage with classroom experience raises questions regarding generalizability. There is also something of a circular argument here, since the use of alternative assessments is expected to help bring more effective instruction in the form of higher order skills and authentic learning into the classroom.

In summary, if LEP students are to be included in alternative assessment programs, then it is important that LEP students as well as mainstream students participate in effective instructional approaches that promote active participation in tasks that require application of higher order learning and problem-solving skills.

3. What are the implications for language minority students?

For LEP students, the shift toward alternative assessment, particularly if the assessment is to be carried out in English, may place an increased burden in terms of the level of language skills required. For example, Secada (1994) suggests that the shift toward use of alternative assessments within mathematics may bring about an increased language load with the end result of greater divergence in mathematics results.

Clearly, for LEP students, the question of the language to be used in assessment is a key one. The advantage of alternative assessment approaches is that, in contrast to standardized tests, alternative assessment tasks can be developed by the teacher in the language of instruction. If portfolio assessment is used, then the student work samples would be in the language in which the student is working. On the other hand, if alternative assessment approaches are to be developed and specific tasks or "item banks" created for the purpose of large-scale use across districts, states, or nationally, then the issue of language and culture as factors in the development of the tasks would clearly remain a difficult technical issue.

4. What are requirements for professional development?

Maldaus (1994) cautions that "It is much easier to plan and mandate a new assessment program than it is to change attitudes and the institutional practices of people involved in the daily life of classrooms" (p. 89). Baron (1992) in describing prerequisites for the effective use of performance-base assessments emphasizes that it is not enough to define new types of assessments. Based on her experience in Connecticut with implementation of performance assessments, she lists seven recommendations which focus on support for teachers, especially those that allow teachers as professionals the time to develop understanding of and participation in the definition of assessment tasks. She also lists administration support and the appropriate curriculum materials that are consistent with the assessments as important elements of implementation.

Teacher educators are beginning to recognize the important role of assessment, and states are beginning to include assessment coursework within requirements for certification or licensure (Cizek, 1993). More than simply additional coursework, however, alternative assessment requires a major conceptual shift on the part of teachers in terms of how they assess and how they instruct, and the relationship between the two. This is not an easy thing to accomplish and much more is needed if alternative assessment is to become the favored approach. In particular, for teachers of LEP students, an understanding of the reason for their students' participation in any assessment, an assurance that results will not be inappropriately used, and an understanding of the specific purposes of the assessment will be important. In particular, this will be necessary in order to obtain their support for inclusion of LEP students in large-scale testing for accountability purposes.

E. Future Directions

Alternative assessment approaches show much promise as a means of assessing students that is consistent with what is considered to be effective instruction. However, the use of alternative assessment must be examined with some caution and care in defining the tasks, the criteria, and the content. As Baker (1992) has noted, there is much more effort needed in definition of alternative assessments, and in ensuring validity and reliability. As a first step, the types of skills to be assessed must be defined, and the interaction of the specific assessment format with the content tested must be examined (Baker, 1992). The purpose for which the assessment will be used must be considered in conjunction with the approach to assessment that is selected, the criteria to be used in making judgements about a student's performance must be identified, and judges trained.

For low stakes purposes or to guide instruction, alternative assessment has shown promise in much of the research. With further definition of tasks, criteria, and scoring, alternative assessment tasks will presumably gain in reliability and validity and in value to teachers and students. However, much of this requires a significant change in beliefs and practices regarding assessment. In addition, for assessment of LEP students, it may require teachers--especially mainstream teachers of LEP students--to shift their beliefs and practices regarding the involvement of LEP students in their classes. These two observations underline the urgent need for a central focus to be placed on professional development issues, and to ensure that there is input from teachers, including teachers of LEP students, incorporated into all efforts to develop new assessment policies and practices.

For use of alternative assessment, expense and time are problems. Even if psychometric worth is in question, however, there is the value of the one-on-one time spent by a teacher and student reviewing a portfolio, or in reviewing progress on a performance task (Maeroff, 1991, in Cizek, 1993). This reflects the validity of alternative assessment's use in the classroom in support of teacher-student partnership in learning that can be useful for the purpose of guiding instruction. Without the necessary reliability and validity established for specific tasks or portfolios, however, then there is no assessment occurring that can be effectively interpreted, generalized, and shared in comparisons or used with confidence in guiding

educational decisions.

F. Summary

In putting alternative assessment approaches into place, the research and practice on use of alternative assessment thus far has indicated the following requirements:

1. Substantial teacher involvement and professional development.
 - Clear definition of the skills to be assessed and of the relationship of the skills to the specific context or format in which they will be assessed.
 - Clear specification of the criteria on which skills are to be judged.
 - Clear specification of the purposes for which an assessment will be used and how the outcomes obtained will be used for that purpose.
 - Specification of scoring and adequate training of raters, especially if assessment is to be used for high stakes decision-making.

For LEP students in particular, alternative assessments do not in themselves resolve some of the key issues, and these will remain to be addressed. For example, in any form of assessment ultimately used, there will still be the question of the language of assessment to be used with LEP students and how to determine this in relation to the student's instructional program and level of language proficiency.

IX. CONCLUSIONS AND RECOMMENDATIONS

As described in the introduction, the goal of this paper was to examine current assessment practices as they relate to LEP students and to suggest some future directions for LEP assessment. The review was intended: (1) to inform federal and state policy-makers as they design standards and systems to assess progress toward national education goals; and (2) to provide suggestions for improving assessment practices at the local level. The review suggests that at both of these levels, it will be necessary to think carefully about the purposes of the assessments to be carried out, the means by which they will be implemented, and the meaning of the scores or other information to be obtained.

There are three broad categories of purposes of assessment which need to be examined: assessment to guide instruction in the classroom; assessment to make decisions regarding selection and placement; and assessment to examine student performance for purposes of school, district, and program accountability. As the discussions above concerning standardized, multiple-choice achievement tests and alternative assessment approaches have indicated, there is not a simple formula to follow. While there is hope that alternative assessments might be designed to eventually address all of these purposes (e.g., Baker, 1992), there are many questions which first need to be asked. While issues related to the definition and use of performance assessment are being examined, the use of performance assessments in conjunction with other more traditional forms of assessment is probably the best route.

Based on the findings of this review of assessment issues and practices relevant to LEP students, the following analyses and recommendations are offered:

A. Assessment in General

Discussion

This review has suggested that assessment is carried out for a variety of purposes, and that the optimal methods of assessment vary based on those purposes. For example, assessments designed for program evaluation and accountability purposes are often not very useful for instructional assessment purposes. However, there is considerable pressure within classrooms and schools to minimize the amount of testing of LEP (and other) students. Thus, specific assessments are often performed for multiple purposes. As a result, assessment results are sometimes expected to be used for purposes for which they were not designed.

Recommendations

1. The purpose or purposes of particular assessments of LEP students should be clearly defined. The assessment should be specifically designed for the purpose(s) for which it is intended. When assessment information is used for purposes other than those for which the assessment was designed, the analysts should clearly state the limitations of the data for addressing those purposes.
- Whenever possible, a particular assessment session should be designed to serve only one purpose. In this way, the assessment can be highly focused on that purpose. If a particular assessment session must serve more than one purpose, the assessment should include subparts which specifically address each of the purposes. An achievement testing session for instructional assessment and accountability purposes, for example, might include both standardized and curriculum-specific subtests.
 - Consideration should be given to use of more than one type of assessment. For example, alternative assessment tasks in combination with modified standardized multiple-choice tests (with items designed to elicit higher order thinking skills to the extent possible) could be developed. In determining the combination, costs and benefits and issues of breadth versus depth of coverage will need to be examined.
 - Any decisions regarding assessment should be made with an awareness of actual implementation requirements and what reasonably can be expected of schools, classrooms, and students. If sufficient attention is not given to the implementation requirements for the school, classroom, or student, then the overall assessment design will be in danger of not being carried out as planned.

B. Assessment of LEP Students for Instructional Purposes

Discussion

The review has suggested that too little attention has been paid to the relationship between assessment and the effective instruction of LEP students. These two issues are often seen as separate topics, and thus there are few models which integrate effective instructional approaches with effective approaches to assessment.

Recommendations

1. LEP students should participate fully in assessment for instructional purposes.
- For the purpose of guiding instruction, alternative assessment approaches should be given continued consideration and development. Alternative assessment approaches to be used in guiding instruction should include those focused on content knowledge and its application, and also assessments focused

on language proficiency skills.

- Teachers should be given additional professional development experiences in the use of assessment in general, in the use of alternative assessment approaches, and in the use of these approaches with LEP students as well as with non-LEP students. Teachers will need sustained follow-up to support them in developing their use of newer assessment approaches.

C. Assessment of LEP Students for Identification and Placement Purposes

Discussion

A major issue identified in this review has been the inappropriate use of assessment results to place LEP students in programs which limit their educational and career opportunities. In particular, there has been concern about inappropriate placement of LEP students in special education and remedial content classes, and in programs which do not lead to graduation and postsecondary education.

An approach which most school districts use to address this issue involves assessment using multiple measures. There is general agreement that no single measure should be used for identification and placement purposes, but there is much less agreement about which tests and selection standards should be used for these purposes.

Recommendations

1. All assessment for the purposes of identification and placement of LEP students should involve the use of multiple measures. Scores on achievement tests in English should not be either the primary or only measure used in assigning LEP students to specific service programs; other means of assessment should be used for these purposes.
- To the extent possible, assessment for the purposes of identification and placement of LEP students should include some measure of native language proficiency. This might range from asking the student or parents about native language proficiency to informal assessment of oral proficiency skills by interview to formal assessment using standardized language proficiency instruments.

D. Assessment of LEP Students for Evaluation and Accountability Purposes

Discussion

This review has discussed the issues and problems associated with including or excluding LEP students from assessments designed to measure the effectiveness of classrooms, schools, and districts in meeting local, state, and federal educational goals. There are strong reasons both for including and excluding LEP students from such assessments if they are conducted in English.

There are a number of reasons for including LEP students in such assessments. If LEP students are included, the overall results for a school, district, or state will be more comprehensive. LEP and other students (e.g., those in special education) are often excluded, resulting in mean achievement levels that are higher than they would be with all students included. Given the variation in how LEP students are defined, consistent inclusion of LEP students can lead to more appropriate or fair comparisons of school units. Also, if LEP students are included, there will be more pressure to improve services to those students.

On the other hand, if LEP students are included in such assessments, the overall results can only be

described as **academic achievement using the English language**. Without the use of parallel versions of the test in the native languages of students (which are few and difficult to produce), it is virtually impossible to separate the effects of content area knowledge and knowledge of the English language. Thus a student who has content area knowledge or skills learned through the native language will not be able to demonstrate them through a test that uses only English.

If LEP students are included in such assessments, there is a very serious danger of misinterpretation of results. Analysts may not recognize that both English language proficiency and content knowledge are tested in achievement tests administered to LEP students, and may under-estimate LEP achievement levels. Thus, for example, students in a school which uses a transitional bilingual education approach may be meeting instructional goals in the native language but may not score well on an English language test.

Including LEP students in such assessments may also have other negative consequences. Taking a test in English may be a highly discouraging experience for a student with limited English skills. If teachers are using the native language or are covering course content which is different from the mainstream program, they will correctly perceive that the test is less relevant and appropriate for their students than for other students.

Recommendations

1. LEP students should not be included in standardized achievement testing programs in English (whether these are traditional standardized tests or alternative assessments) until they have received formal language instruction in English for a minimum specified period of time; we recommend one and one-half years.
- When LEP students are included in standardized achievement testing in English, the results should be used **only** for group assessment purposes. Such scores can serve to define how subgroups of LEP students are progressing in relation to instructional objectives as measured in English. To the extent possible, the results should be disaggregated to examine the achievement of meaningful subgroups of LEP students based on length of time in U.S. schools, levels of English language proficiency and, especially, type of service being received.
- Whenever results are presented involving standardized achievement test scores in English for LEP students, the presentation should clearly indicate that achievement test outcomes for LEP students are the outcome of two sources of input: English language proficiency and academic knowledge. The presentation should also state that academic achievement levels of LEP in English may not be related to their achievement levels in the native language.
- Considerable work needs to be done in developing a national consensus concerning how to report results for LEP students included in standardized achievement testing in English. The results for LEP students could be presented separately or as part of overall school/district means.
- For LEP students in bilingual programs, achievement testing for content subjects should if possible be performed in the language in which the student is most capable of showing knowledge and skills. The results of standardized achievement testing in the native language can be used for instructional assessment, placement, and program evaluation purposes.
- When achievement testing of LEP students is being done for program evaluation or accountability purposes only, consideration should be given to testing at fewer grade levels or to sampling students

within grade levels. The testing burden on LEP students is very high, and such methods could decrease the testing burden while still providing for the necessary inclusion of LEP students in analyses of performance.

APPENDIX A:

References

- Abbott, M. (1985). *Theoretical considerations in the measurement of English language proficiency of limited English proficient students*. Presentation at the annual meeting of the National Council on Measurement in Education.
- Airasian, P.W. (1987). State mandated testing and educational reform: Context and consequences. *American Journal of Education*, May, 393-412.
- Anderson, R., Hiebert, E.H., Scott, J.A., and Wilkinson, I.A.G. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington, DC: National Institute of Education.
- Ascher, C. (1990). *Assessing bilingual students for placement and instruction*. ERIC Clearinghouse on Urban Education Digest, Number 65.
- Bachman, L.F. (1990). Assessment and evaluation. *Annual Review of Applied Linguistics*. USA: Cambridge UP, 210-226.
- Baker, E.L. (1992). Issues in policy, assessment, and equity. In *Proceedings of the second national research symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 2*. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Ballard, W.S., Tighe, P.L., and Dalton, E.F. (1991). *Idea Oral Language Proficiency Test*, English edition. Brea, CA: Ballard & Tighe.
- Baron, J.B. (1989). Performance testing in Connecticut. *Educational Leadership*, 46(7), 8.
- Baron, J. (1992). SEA usage of alternative assessment: The Connecticut experience. In *Proceedings of the second national research symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 2*. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Bermudez, A.B. and Rakow, S.J. (1990). Analyzing teachers' perceptions of identification procedures for gifted and talented Hispanic limited English proficient (LEP) students at risk. *The Journal of Educational Issues of Language Minority Students*, 7, 21-33.
- Board of Education, City of New York Office of Testing, Division of Curriculum and Instruction. (1982). *Language Assessment Battery*, English edition. New York, NY: New York City Board of Education.
- Brainard, C. (1978). Review of Language Assessment Scales, LAS I for grades K-5, English/Spanish. *Modern Language Journal*, 62(7), 350-1.

- Brown, A.L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., and Campione, J.C. (1991). Distributed expertise in the classroom. In G. Salomon (Ed.) *Distributed cognitions*. New York: Cambridge University Press.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, Mass.: Harvard University Press.
- Burt, M.K., Dulay, H.C., Chavez, E. H. (1978). *Bilingual Syntax Measure I and II*, English edition. San Antonio, TX: The Psychological Corporation.
- California Achievement Tests*, level 14, form A. Fifth edition. (1992). Monterrey, CA: CTB Macmillan/McGraw-Hill.
- California Achievement Tests*, technical bulletin 1. Fifth edition. (1992). Monterrey, CA: CTB Macmillan/McGraw-Hill.
- Canales, J. (1994). *Alternative assessments: Measuring the linguistic, cognitive, and affective/social performance of language minority students*. Presentation at the National Association for Bilingual Education Annual Conference, OBEMLA Professional Development Institute. February. Los Angeles, CA.
- Casas, J. and Strand, D. (1989). Maculaitis Assessment Program, commercial edition. In J. Conoley, J. Kramer (Eds.) *The tenth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Chesterfield, K.B. (1985). Review of Language Assessment Battery. In J.V. Mitchell Jr. (Ed.) *The ninth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Cizek, G. (1993). Some thoughts on educational testing: Measurement policy issues into the next millennium. *Educational Measurement: Issues and Practice*, 12(3) 10-16, 22.
- Collier, V. (1992). A synthesis of studies examining long-term language minority study data on academic achievement. *Bilingual Education Research Journal*, 16(1/2), 187-221.
- Comprehensive Tests of Basic Skills*, level 14, form A. Fourth edition. (1989). Monterrey, CA: CTB/McGraw-Hill.
- Comprehensive Tests of Basic Skills*, technical report. Fourth edition. (1989). Monterrey, CA: CTB/McGraw-Hill.
- Council of Chief State School Officers. (1992). *Recommendations for improving the assessment and monitoring of students with limited English proficiency*. Washington, DC: Author.
- Cummins, J. (1984). Empowering language minority students: A framework for intervention. *Harvard Educational Review*, 56, 18-35.
- Cziko, G.A. (1987). Review of the Bilingual Syntax Measure I. In J.C. Alderson, K.J. Krahnke, and C.W. Stansfield (Eds.) *Reviews of English language proficiency tests*. Washington, DC: Teachers of English to Speakers of Other Languages.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.

- DeAvila, E. (1994). *Recommendations for educational assessment concerning language-minority students*. Presentation at National Association for Bilingual Education, conference. Los Angeles, CA.
- DeAvila, E. (1990). *Assessment of language minority students: Political/technical/Practical and moral imperatives*. Presentation at National Symposium on Limited English Proficient Students' Research Issues. Washington, DC.
- DeAvila, E.A. and Duncan, S.E. (1990). *Language Assessment Scales*. English edition. Monterrey, CA: CTB Macmillan McGraw-Hill. 1990.
- Dunn, Lloyd M., and Dunn, Leota M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- Feuer, M.J. and Fulton, K. (1993). The many faces of performance assessment. *Phi Delta Kappan*, February, 478.
- Fleischman, H.L. and Hopstock, P.J. (1993). *Descriptive study of services to limited English proficient students. Volume 1: Summary of findings and conclusions*. Report to the U.S. Department of Education, Office of the Under Secretary. Arlington, VA: Development Associates, Inc.
- Gandara, P. and Merino, B. (1993). Measuring the outcomes of LEP programs: Test scores, exit rates and other mythological data. *Educational Evaluation and Policy Analysis*, 22(3), 320-338.
- Garcia, E. (1989). Maculaitis Assessment Program, commercial edition. In J. Conoley, and J. Kramer (Eds.) *The tenth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Garcia, E. (1985). Review of Bilingual Syntax Measure II. In J.V. Mitchell, Jr. (Ed.) *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements & the University of Nebraska.
- Garcia, G.N. (1994). Bilingual education: A look to the year 2000. *Focus: Occasional Papers in Bilingual Education*, 9, Winter/Spring. Washington, DC: National Clearinghouse for Bilingual Education.
- Geisinger, K. (1992). Testing limited English proficient students for minimum competency and high school graduation. In *Proceedings of the second national research symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 2*. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Gillmore, G., and Dickerson, A. (1979). *The relationship between instruments used for identifying children of limited English-speaking ability in Texas*. Arlington, VA: Education Resources Information Center. (ERIC Document Reproduction Service No. ED 191907).
- Haber, L. (1985). Review of Language Assessment Scales. In J.V. Mitchell Jr. (Ed.) *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements & the University of Nebraska.
- Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. New York: Basic Books, Inc.
- Hamayan, E. and Damico, J. (1991). *Limiting bias in the assessment of bilingual students*. Austin, TX: Pro-ed.

- Haney, W. and Madus, G. (1989). Searching for alternatives to standardized tests: whys, whats, and whithers. *Phi Delta Kappan*, May, 683-687.
- Hayes-Brown, Zoe Ann. (1987). Review of the Bilingual Syntax Measure II. In J.C. Alderson, K.J. Krahnke, and C.W. Stansfield (Eds.) *Reviews of English language proficiency tests*. Washington, DC: Teachers of English to Speakers of Other Languages.
- Hopstock, P.J., Bucaro, B.J., Fleischman, H.L., Zehler, A.M., and Eu, H. (1993). *Descriptive study of services to limited English proficient students. Volume II: Survey results*. Report to the U.S. Department of Education, Office of Policy and Planning. Arlington, VA: Development Associates, Inc.
- Iowa Tests of Basic Skills*, level 9, form K. (1993). Chicago, IL: Riverside Publishing Company.
- Iowa Tests of Basic Skills*, technical summary 1. (1994). Chicago, IL: Riverside Publishing Company.
- Koretz et al. (1993). *The reliability of Vermont portfolio scores in the 1992-1993 school years*. (Interim Report.) CSE Technical Report. Washington DC: Rand Institute.
- LaCelle-Peterson, M. and Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review*, 64(1), 55-75.
- Lam, T.C. (1988). *Testing, opportunity allocation, and Asian and pacific Americans*. Report prepared for the National Commission on Testing and Public Policy. Berkeley, CA: Graduate School of Education.
- Maculaitis, J. D. (1991). *The Maculaitis Assessment Program: Multi-purpose tests for non-native speakers of English*. Englewood Cliffs, NJ: Alemany Press/Prentice Hall Regents.
- Madaus, G. (1989). The influence of testing on the curriculum. In L.N. Tanner (Ed.) *Critical issues in curriculum: 87th yearbook of the National Society for the Study of Education*. Chicago, IL: University of Chicago Press.
- Madaus, G.F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Educational Review*, 64(1), 76-95.
- Maeroff, G.I. (1991). Assessing alternative assessment. *Phi Delta Kappan*, December, 273- 281.
- McCullum, P.A. (1983). The Idea Oral Language Proficiency test: A critical review. In S.S. Seidner (Ed.) *Issues of language assessment. Volume II: Language assessment and curriculum planning*. Chicago: Illinois State Board of Education.
- McCullum P.A. (1987). Review of the Idea Oral Language Proficiency Test. In J.C. Alderson, K.J. Krahnke, and C.W. Stansfield (Eds.) *Reviews of English language proficiency tests*. Washington, DC: Teachers of English to Speakers of Other Languages.
- McLaughlin, B. (1987). *Theories of second language learning*. London: Edward Arnold.
- Measuring progress toward America's education goals*. Stanford Achievement Test scores, eighth edition. (1992). San Antonio, TX: Psychological Corporation.

Mehrens, W.A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11(1), 3-9, 20.

Meisels, S.J. (1989). High-stakes testing in kindergarten. *Educational Leadership*, 46(7), 16-21.

Metropolitan Achievement Test, elementary 1, form S. Seventh edition. (1993). San Antonio, TX: Psychological Corporation.

Metropolitan Achievement Test, technical data report. Eighth edition. (1993). San Antonio, TX: Psychological Corporation.

Milk, R. (1993). Bilingual education and English as a second language: The elementary school. In M.B. Arias and U. Casanova (Eds.) *Bilingual Education: Politics, Practice and Research*. Ninety-Second Yearbook of the National Society for the Study of Education. Chicago, IL: University of Chicago Press.

Miller, M.D. and Legg, S.M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12(2), 9-15.

Minicucci, C. and Olsen, L. (1992). Programs for secondary limited English proficient students: A California study. *Focus: Occasional Papers in Bilingual Education*, 5, Spring. Washington, DC: National Clearinghouse on Bilingual Education.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: Government Printing Office.

Navarrete, C., Wilde, J., Nelson, C., Martinez, R. and Hargett, G. (1990). *Informal assessment in educational evaluation: Implications for bilingual education programs*. National Clearinghouse for Bilingual Education: Program information guide series 3, summer. Washington, DC: The George Washington University.

Newmann, F.M. and Wehlage, G.G. (1993). Five standards of authentic instruction. *Educational Leadership*, 50(7), 1-12.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.

O'Brien, M.L. (1989). Psychometric issues relevant to selecting items and assembling parallel forms of language proficiency instruments. *Educational and psychological measurement*, 49, 347-353.

Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions*. Washington, DC: Author.

Oller, J. (1992). Language testing research: Lessons applied to LEP students and programs. In *Proceedings of the second national research symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 2*. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs.

Oller, J. (1979). *Language tests at school*. London, England: Longman Group Limited.

- Oller, J.W. (1976). Review of the Bilingual Syntax Measure. *Modern Language Journal*, 60(7), 399-400.
- Ortiz, A.A. (1992). Assessing appropriate and inappropriate referral systems for LEP special education students. In *Proceedings of the Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 1*. Washington, DC: United States Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Ortiz, A. and Wilkinson, C. (1990). Assessment and intervention model for the bilingual exceptional student. *Teacher Education and Special Education*, 35-42.
- Peterson, P.L. (1992). Response to Walter Secada's presentation. In *Proceedings of the Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 2*. Washington, DC: United States Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Resnick, L.B. (1987). *Education and learning to think*. Oxford: Oxford University Press.
- Rivera, C. and Zehler, A.M. (1991). Assuring the academic success of language minority students: Collaboration in teaching and learning. *Journal of Education*, 173(2), 52-77.
- Rodriguez-Brown, F. and Elias-Olivares, L. (1981). *A search for congruency in language proficiency testing: What the tests measure -- what the child does*. Presentation at the annual meeting of the American Educational Research Association.
- Rosansky, E.J. (1979). A Review of the Bilingual Syntax Measure. In B. Spolsky (Ed.) *Some major tests. Advances in Language Testing Series: 1*. Arlington, VA: Center for Applied Linguistics.
- Rueda, R. (1992). Response to Joan Canales' presentation. In *Proceedings of the Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement. Volume 2*. Washington, DC: United States Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Ruiz-Primo, M.A., Baxter, G.P., and Shavelson, R.J. (1993). On the stability of performance assessments. *Journal of Educational Measurement*, 30(1), 41-53.
- Schiff-Myers, N.B., Djukic, J., McGovern-Lawler, J., and Perez, D. (1993). Assessment considerations in the evaluation of second-language learners: A case study. *Exceptional Children*, 60(3), 237-248.
- Secada, W. (1994). Issues in the development of Spanish-language versions of the national assessment of educational progress. In *The trial state assessment: Prospects and realities: Background studies*. Stanford, CA: The National Academy of Education.
- Shavelson, R.J., Baxter, G.P., and Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215-232.
- Shavelson, R.J., Baxter, G.P., and Pine, J. (1991). Performance assessment in science. *Applied measurement in education*. 4(4), 347-362.
- Shellenberger, S. (1985). Review of Bilingual Syntax Measure II. In J.V. Mitchell Jr. (Ed.) *The ninth mental*

- measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements & the University of Nebraska.
- Shepard, L.A. (1989). Why we need better assessment. *Educational Leadership*, 46(7), 4-9.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. Baltimore, MD: National Foreign Language Center.
- Short, D.J. (1993). Assessing integrated language and content instruction. *TESOL Quarterly*, 27(4), 627-656.
- Sosa, Alicia, S. (1990). *Assessment of language minority students*. Regional Hearing - Education of Hispanics. San Antonio, TX: Intercultural Development Research Center.
- Special Issues Analysis Center. (1993). *Active learning instructional models for limited English proficient (LEP) students*. Focus group report. Arlington, VA: Development Associates, Inc.
- Spencer, B.D. (1994). A study of eligibility exclusions and sampling: 1992 trial state assessment. In *The trail state assessment: Prospects and realities: Background studies*. Stanford, CA: The National Academy of Education.
- Stanford Achievement Test*, primary 3, form L. Eighth edition. (1992). San Antonio, TX: Psychological Corporation.
- Stanford Achievement Test*, technical data report. Eighth edition. (1990). San Antonio, TX: Psychological Corporation.
- Ulibarri, D.M., Spencer, M. and Rivas, G.M. (1981). Language proficiency and academic achievement: A study of language proficiency tests and their relationship to school ratings as predictors of academic achievement. *NABE Journal*, 5(3), 47-80.
- Valdez-Pierce, L., and O'Malley, J.M. (1992). *Performance and portfolio assessment for language minority students*. National Clearinghouse on Bilingual Information: Program Information Guide Series, 9. Summer.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.
- Waggoner, D. (1994). Language minority households number more than 14 million. *Numbers and Needs*, 4(1), 1-2.
- Warren, B. and Rosebery, A. (1990). *Cheche Konnen: Collaborative scientific inquiry in language minority classrooms*. Technical Report from the Innovative Approaches Research Project. Arlington, VA: Development Associates, Inc.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Worthen, B.R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Beta Kappa*, February. 444-456.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Walter Secada and Dr. Judy Torres who reviewed and offered comments on an earlier version of this report. We also wish to thank Ms. Carolyn Vincent for her suggestions and acknowledge the assistance of Ms. Jean Bender in the initial reviews of the assessment instruments. While the report would have been impossible without the support of these individuals, the authors take full responsibility for the opinions, conclusions and recommendations expressed herein.

This report was prepared for the U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs, under Contract No. T292001001, Task Order No. D070. The opinions, conclusions, and recommendations expressed herein do not necessarily reflect the position or policy of the Department of Education and no official endorsement by the Department of Education should be inferred.

[go to HOME PAGE](#)

www.ncela.gwu.edu